حل تمرین social

سوال اول

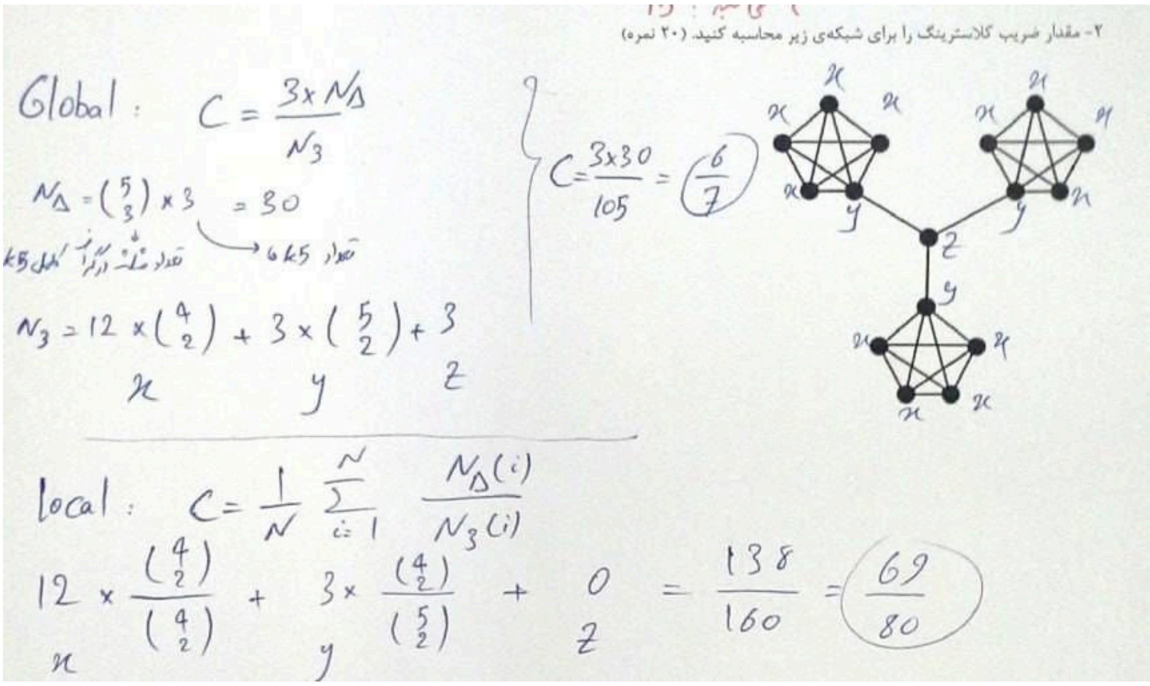
مقدار ضریب کلاسترینگ را برای شبکه زیر محاسبه کنید.

پاسخ:

برای یک راس: تعداد یال های بین همسایه ها تقسیم بر کل یال های ممکن بین آن ها

گلوبال (جدید) : تعداد کل مثلث ها تقسیم بر تعداد کل گیلاس ها ضرب در ۳

لوکال (فرمول اسلاید): میانگین ضریب کلاسترینگ رئوس



سوال دوم

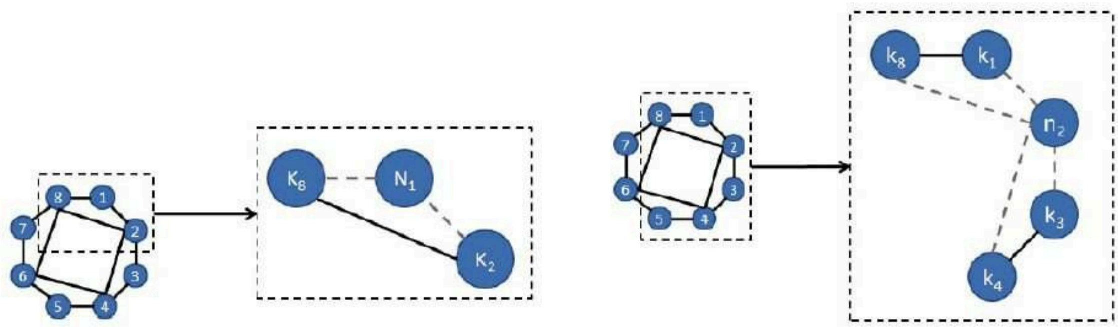
یک حلقه زوج رأسی داریم. علاوه بر یال های حلقه، رأس های زوج به رأس زوج قبل و بعد از خود نیز متصل هستند. میانگین ضریب کلاستریگ چقدر است؟

پاسخ

نود های فرد ۲ همسایه دارند که به هم متصل اند. پس ضریب کلاستریگ برای آن ها ۱ میشود.

نود های زوج ۴ همسایه دارند و دو یال بین همسایه های آن ها موجود است. ضریب کلاسترینگ برای آن ها ۲/۶ میشود.

در نتیجه میانگین ضریب کلاسترینگ ۲/۳ است. (تعداد رأس های زوج و فرد برابر است)



odd nodes even nodes

All the odd nodes will have a clustering coefficient of 1 because they only have two neighbors and those two neighbors know one another.

The even nodes have four neighbors, and the two pairs of neighbors on either side know one another. That is 2 edges out of $\binom{4}{2=6}$ possible ones. So their clustering coefficient is $2/6 = 1/3$.

The average clustering coefficient for the whole network is therefore

$$\frac{1}{2}(1+\frac{1}{3}) = \frac{2}{3} \tag{1}$$

سوال سوم

توزیع stationary قدم زدن تصادفی (در صورت وجود) برای یک گراف با ماتریس مجاورت A چگونه به‌دست می آید؟

پاسخ:

تعریف رفتار stationary: بعد ازبه اندازه کافی گذر در رئوس احتمال حضور در رأس ها تغییر نکند.

ماتریس P از روی ماتریس A بدست می آید. به این صورت که احتمال رفتن از i به j برای همسایه های i میشود ۱ بر روی درجه. برای باقی رئوس صفر

اگر X_t توزیع احتمال حضور در راس های گراف در مرحله t را نشان دهد، داریم

$$X_{t+1} = X_t P$$

اگر برای t به اندازه کافی بزرگ $X_{t+1} = X_t$ انگاه X_t یک توزیع stationary است.

بردار ویژه متناظر با مقدار ویژه 1 جواب مسئله است

V = VP

سوال چهارم

بررسی page rank شخصی سازی شده

در این روش کاربران مختلف پیج های مرتبط متفاوتی میبینند. موتور جست و جو بر اساس علایق کاربر به آن ها پیشنهاد میدهد. میتوان علایق را با teleport set مدل کرد.

فرض کنید علایق کاربران را با یک مجموعه از اسامی سایت ها نشان دهیم. برای سادگی از عدد برای نمایش اسم سایت استفاده میکنیم. به عنوان مثال علاقه به اسنپ فود و ورزش ۳ را با مجموعه {2, 5}

فرض کنید بردار پیج رنک را برای کاربران زیر داریم

- آناهیتا {1, 2, 3}
- بهروز {3, 4, 5}
- کامران {1, 4, 5}
- درسا {1}

الف) بدون داشتن گراف، آیا میتوانید برای کاربران زیر بردار پیج رنک شخصی سازی شده را بدست آورید؟ اگر نه چرا؟ اگر بله چطور؟ پارامتر تلپورت را بتا بگیرید

- علی با بردار علاقه مندی {2}
- فرید با بردار علاقه مندی {5}
- قاسم با بردار علاقه مندی {1, 2, 3, 4, 5} و وزن های 0.1, 0.2, 0.3, 0.2, 0.2

ب) فرض کنید بردار های برخی کاربران محاسبه شده اند. مجموعه همه بردار های شخصی سازی شده ای که میتوان بدون داشتن گراف یافت کدام است؟

پاسخ:

ابتدا ثابت میکنیم اگر این مجموعه در اسپم مجموعه های موجود باشد، میتوان بردار آن را با ترکیب خطی بردار های موجود بدست آورد.

پاسخ ب:

همه ترکیب خطی های V که میشود span(V)

(a) [**7 points**] Suppose you have already computed the personalized PageRank vectors for the following users:

- Agatha, whose interests are represented by the teleport set $\{1, 2, 3\}$,
- Bertha, whose interests are represented by the teleport set $\{3, 4, 5\}$,
- Clementine, whose interests are represented by the teleport set $\{1, 4, 5\}$, and
- DeShawn, whose interests are represented by the teleport set $\{1\}$.

Without looking at the graph, can you compute the personalized PageRank vectors for the following users? If so, how? If not, why not? Assume a fixed teleport parameter β.

★ **Solution** ★ We can compute the personalized PageRank vectors for any user whose interest vector is spanned by the interest vectors for the users we already computed. For simplicity of notation we will show this for two users whose interests are represented by the vectors s_A and s_B and their personalized PageRank vectors are r_A and r_B. Now assume that $s_C = qs_A + (1-q)s_B$. Then the personalized PageRank vector for user C is $r_C = qr_A + (1-q)r_B$, because:

$$\begin{aligned} M'_C r_C =& \beta M r_C + (1-\beta)\, s_C \\ =& q\left(\beta M r_A + (1-\beta)\, s_A\right) + (1-q)\left(\beta M R_B + (1-\beta)\, s_B\right) \\ =& q r_A + (1-q) r_B = r_C \end{aligned}$$

i. [**2 points**] Eloise, whose interests are represented by the teleport set $\{2\}$.

★ **Solution** ★ Yes. $3A - 3B + 3C - 2D$.

ii. [**2 points**] Felicity, whose interests are represented by the teleport set $\{5\}$.

★ **Solution** ★ No. There is no way to isolate 5 without involving B, which in turn involves A, and there is no way to cancel out the resultant 2.

iii. [**3 points**] Glynnis, whose interests are represented by the teleport set $\{1, 2, 3, 4, 5\}$ with weights 0.1, 0.2, 0.3, 0.2, 0.2, respectively.

★ **Solution** ★ Yes. $.6A + .3B + .3C - .2D$

(b) [**3 points**] Suppose that you've already computed the personalized PageRank vectors of a set of users (denote the computed vectors V). What is the set of all personalized PageRank vectors that you can compute from V without accessing the web graph?

در تصویر بالا M احتمال انتقال بین سایت ها است.
از روی مجموعه های علاقه مندی بردار s ساخته میشود که توزیع احتمال عمل کردن بر اساس علاقه مندی است. در این حالت فرد به جای عمل کردن بر اساس ماتریس M به سایت های مورد علاقه اش با احتمال یکسال میرود.
در نهایت r بردار شخصی سازی شده علاقه مندی است.
اگر بردار Sc از روی Sa, Sb بدست آید و rc را ترکیب خطی ra و rb قرار دهیم در رابطه پیج رنک صدق میکند.

سوال پنجم

توضیح: excess degree distribution توزیع احتمال $\{q_k\}$ است که نشان میدهد یک یال رندم با چه احتمالی در یک رأس با درجه $k+1$ رفته است. برای محاسبه آن میتوان تعداد چنین یال هایی را برای هر k شمارش کرد و توزیع احتمالی با همین نسبت ایجاد کرد. برای این منظور (q'_k) $q_k = q'_k/2E$ بدست می آید.
امید ریاضی این توزیع نیز مشابه توزیع درجات محاسبه میشود.

An important concept in network analysis is the *excess degree distribution*, denoted as q_k, for $k \geq 0$. Intuitively, q_k gives the probability that a randomly chosen edge goes to a node of degree $k+1$. Excess degree can be calculated as follows:

$$q_k = \frac{q'_k}{\sum_i q'_i}, \qquad q'_k = \sum_{i \in V} \sum_{(i,j) \in E} I_{[k_j = k+1]},$$

where $I_{\text{condition}} = 1$ when condition is true and 0 otherwise. V denotes the set of nodes, E the set of edges and k_j the number of neighbors of node j (equivalently, the degree of node j). Additionally, the *expected excess degree* is defined as $\sum_{k \geq 0} k \cdot q_k$. The degree distribution of the network is denoted by $\{p_k | k \geq 0\}$, i.e., p_k is the proportion of nodes having degree exactly k. The *expected degree* can consequently be computed as $\sum_{k \geq 0} k \cdot p_k$.

الف) برای یک گراف Expected degree , Expected excess degree را محاسبه کنید.

ب) اگر توزیع درجات به صورت $\{p_k\}$ در دسترس باشد، آیا میتوان توزیع excess degree را با استفاده از آن بدست آورد؟ فرمولی برای این کار ارائه دهید.

پاسخ

الف)

امید ریاضی درجه میشود تعداد یال ضرب در ۲ تقسیم بر تعداد رئوس. $E(p_k) = 2m/n$

برای محاسبه مقدار Expected excess degree، کافی است امید ریاضی q'_k را محاسبه کنیم.

به این توجه کنید که برای هر یال و راس، درجه راس منهای یک در مقدار امید ریاضی محاسبه شده است. زیرا در رابطه $\sum kq'_k$ مقدار k در تعداد یال هایی ضرب شده است که راسشان درجه k+1 دارد.

هر راس به اندازه درجه اش یال دارد. پس برای هر راس $d(d-1)$ مقداری است که در امید ریاضی محاسبه میشود. در نتیجه امید ریاضی از رابطه زیر بدست می آید

$E(q'_k) = \sum_i d_i(d_i - 1)$

در نتیجه

$E(q_k) = (\sum_i d_i^2 - 2m)/2m$

ب)

- **(ii) [5 points]** Can you calculate the excess degree distribution $\{q_k\}$ using the degree distribution $\{p_k\}$? Find this formula.

★ Solution ★

- q_k is the probability that a random edge goes to a node of degree $k+1$, calculated as the number of edges that go to a node of degree $k+1$ over the total number of edges.
- The number of edges that go to a node of degree $k+1$ is $n(k+1)p_{k+1}$.
- Since the expected degree is $\Sigma_k k p_k$, the total number of edges is $n \Sigma_k k p_k$.
- So, $q_k = \frac{n(k+1)p_{k+1}}{n\Sigma_k k p_k}$, or after canceling n, $\boxed{q_k = \frac{(k+1)p_{k+1}}{\Sigma_k k p_k}}$